IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants	:	SAITOU, et al.
Serial No.	:	10/646,390
Filed	:	August 21, 2003
Examiner	:	TBA
Group Art Unit	:	TBA
For	:	**GENES**

745 Fifth Avenue
New York, New York 10151

EXPRESS MAIL
Mailing Label Number: EV 287822612US
Date of Deposit: January 9, 2004
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" Service under 37 CFR 1.10 on the date indicated above and is addressed to: **Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Saddam Ahmed
(Typed or printed name of person mailing paper or fee)

(Signature of person mailing paper or fee)

CLAIM OF PRIORITY

Commissioner for Patent
Alexandria, VA 22313-1450

Sir:

Applicants hereby claim priority under 35 U.S.C. §119 and/or 120, from UK patent application number 0101300.2 and International patent application number PCT/GB02/00215, a certified copy of each is enclosed.

Acknowledgment of the claim of priority and of the receipt of said certified copy are respectfully requested.

Respectfully submitted,

FROMMER LAWRENCE & HAUG LLP

By: Thomas J. Kowalski
THOMAS J. KOWALSKI, Reg. No. 32,147
(212) 588-0800

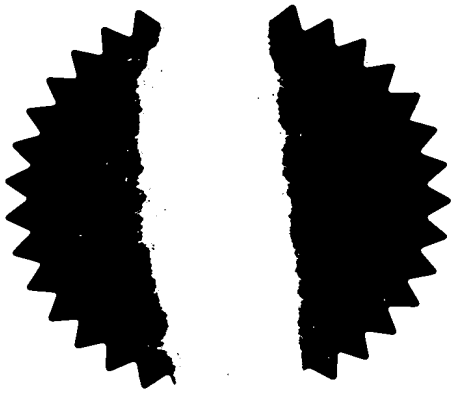
The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed Andrew Gerry

Dated 27 November 2003

Patents Form 1/77

Patent Office

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
A
18 JAN 2001
RECEIVED BY FAX

Request for a grant of a patent

The Patent Offic

Cardiff Road
Newport
Gwent NP9 1RH

(See the notes on the back of this form you can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference — P010490GB ATM

2. Patent application number *(The Patent Office will fill in this part)* — **0101300.2** — 18 JAN 2001

P01/7700 0.00-0101300.2

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

Cambridge University Technical Services Limited
20 Trumpington Street
Cambridge
CB2 1QA
United Kingdom

18JAN01 E599117-1 D02246
P01/7700 0.00-0101300.2

Patents ADP number *(if you know it)* — 7882871003

If the applicant is a corporate body, give the country/state of its incorporation — United Kingdom

4. Title of the invention — PRIMORDIAL GERM CELL GENES

5. Name of your agent *(if you have one)* — **D YOUNG & CO**

"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

21 NEW FETTER LANE
LONDON
EC4A 1DA

Patents ADP number *(if you know it)* — **59006**

6. If you are declaring priority from one or more earlier patent applications, give the country and date of filing of the or each of these earlier applications and *(if you know it)* the or each application number

	Country	Priority application number *(if you know it)*	Date of filing *(day/month/year)*
1st			
2nd			
3rd			

7. If this application is divided or otherwise derived from an earlier UK application, give the number and filing date of the earlier application

Number of earlier application	Date of filing (day/month/year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

YES

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	31 ✓
Claim(s)	2 ✓
Abstract	1 ✓
Drawing(s)	–

10. If you are also filing any of the following, state how many against each item

Priority Documents	–
Translation of Priority Documents	–
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	–
Request for preliminary examination and search *(Patents Form 9/77)*	–
Request for substantive examination *(Patents Form 10/77)*	–
Any other documents *(Please specify)*	SEQUENCE LISTING (2 pages)

11. I/We request the grant of a Patent on the basis of this application.

Signature D. Young & Co

Date 18 Jan 2001

D YOUNG & CO
Agents for the Applicants

12. Name and daytime telephone number of person to contact in the United Kingdom

Antonio Maschio ~~023 80634816~~ 023 8071 9500

Warning
After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes
a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheets should be attached to this form.

d) If you answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

Primordial Germ Cell Genes

The present invention relates to genes which are expressed exclusively in the earliest populations of primordial germ cells (PGCs) and the use of said genes and the products thereof in identification of PGCs in cell populations.

Introduction

Post fertilisation, the early mammalian embryo undergoes four rounds of cleavage to form a morula of 16 cells. These cells, following further rounds of division, develop into a blastocyst in which the cells can be divided into two distinct regions; the inner cell mass, which will form the embryo, and the trophectoderm, which will form extra-embryonic tissue, such as the placenta.

The cells that form part of the embryo up until the formation of the blastocyst are totipotent; in other words, each of the cells has the ability to give rise to a complete individual embryo, and to all the extra-embryonic tissues required for its development. After blastocyst formation, the cells of the inner cell mass are no longer totipotent, but are pluripotent, in that they can give rise to a range of different tissues. A known marker for such cells is the expression of the enzyme alkaline phosphatase.

Primordial germ cells (PGCs) are pluripotent cells that have the ability to differentiate into all three primary germ layers. In mammals, the PGCs migrate from the base of the allantois, through the hindgut epithelium and dorsal mesentery, to colonise the gonadal anlague. The PGC-derived cells have a characteristically low cytoplasm/nucleus ratio, usually with prominent nucleoli. PGCs may be isolated from the embryos by removing the genital ridge of the embryo, dissociating the PGCs from the gonadal anlague, and collecting the PGCs. The earliest PGC population is reported to consist of a cluster of some 40 alkaline phosphatase positive cells, found at the base of the emerging allantois. 7.25 days post-fertilisation (Ginsburg *et al.*, (1990) Development 110:521-528).

PGCs have many applications in modern biotechnology and molecular biology. They are useful in the production of transgenic animals, where embryonic germ (EG) cells derived

from PGCs may be used in much the same manner as embryonic stem (ES) cells (Labosky *et al.*, (1994) Development 120:3197-3204). Moreover, they are useful in the study of foetal development and the provision of pluripotent stem cells for tissue regeneration in the therapy of degenerative diseases and repopulation of damaged tissue following trauma. However, such cells are difficult to isolate from embryonic tissue, which complicates their study and the development of techniques which make use thereof.

Little is known in the art about the expression of genes in PGCs and the relationship between PGC-specific gene expression and the commitment of a totipotent cell to the germ cell fate. Certain markers for PGCs are known – for example, the expression of tissue non-specific alkaline phosphatase (TNAP) has been used as a marker for early PGCs (Ginsburg *et al.*, (1990) Development 110:521-528). Oct4 is known to be expressed in totipotent PGCs, but not somatic cells (Yoem *et al.*, (1996) Development 122:881-894). Other markers, such as BMP4, are known to be expressed in somatic tissues but not PGCs (Lawson *et al.*, (1999) Genes & Dev. 13:424-436). However, none of these genes is specific for PGCs, since they are also expressed in other tissue types. There is therefore a need in the art for the identification of genes which may be used as markers for PGCs and which may provide an insight into the biology of germ cell development.

Summary of the Invention

The present invention provides two genes which are expressed specifically in PGCs. The sequence of the genes of the invention is set forth in SEQ. ID. No. 1 (GCR1) and SEQ. ID. No. 3 (GCR2). According to a first aspect, therefore, the invention provides a nucleic acid molecule which is at least 90% homologous to SEQ. ID. No. 1 and a nucleic acid molecule which is at least 75% homologous to SEQ. ID. No. 3.

GCR1 and GCR2 are PGC-specific transcripts. GCR1 is upregulated during the process of lineage commitment of PGCs, while GCR2 is upregulated after GCR1, and marks commitment to the PGC fate. GCR1 encodes a 137 amino acid polypeptide of approximately 15kd, which is a transmembrane protein having at least one extracellular

domain. It is a member of a multigene family, and a plurality of members of the family have been isolated from PGCs. The GCR1 polypeptide shows 89% homology to the interferon-inducible protein (sp: INIB RAT, pir:JC1241; GenPept GI:111876). It is considered that the latter polypeptide is the rat homologue of GCR1. GCR2 is a 150 amino acid polypeptide, of approximately 18kd, with a very basic pI of 9.76. The polypeptide comprises a nuclear localisation signal, and is a nuclear protein. It has no homology to any known protein.

The invention further provides polynucleotides which comprise a contiguous stretch of nucleotides from SEQ. ID. No. 1 or SEQ. ID. No. 2, or of a sequence at least 90% homologous thereto. Advantageously, this stretch of contiguous nucleotides is 50 nucleotides in length, preferably 40, 35, 30, 25, 20, 15 or 10 nucleotides in length.

The genes GCR1 and GCR2 encode novel polypeptides, the sequences of which are set forth in SEQ. ID. No. 2 and SEQ. ID. No. 4. The invention therefore encompasses polypeptides encoded by the nucleic acids according to the invention. Preferably, the polypeptides have the sequences set forth in SEQ. ID. No. 2 and SEQ. ID. No. 4.

Antibodies may be raised against the polypeptides according to the invention. In particular, antibodies may be raised against the extracellular domain of GCR1, which is a transmembrane polypeptide.

Antibodies and nucleic acids according to the invention are useful for the identification of PGCs in cell populations. The invention therefore provides a means to isolate PGCs, useful for example for the study of germ tissue development and the generation of transgenic animals, and PGCs when isolated by a method according to the invention.

Homologues of GCR1 and GCR2, such as rat interferon-inducible protein, may also be used to identify PGCs according to the present invention.

Moreover, the invention provides a method by which genes specifically expressed in PGCs may be isolated, comprising the steps of:

a) providing a population of cells containing PGCs;

b) isolating one or more PGCs therefrom and providing single-cell PGC isolates;

c) amplifying the transcribed nucleic acid present in a single PGC;

d) conducting a subtractive hybridisation screen to identify transcripts present in PGCs but not in somatic cells; and

e) probing a nucleic acid library with one or more transcripts identified in d) to clone one or more genes which are specifically expressed in PGCs.

Detailed description of the Invention

Although in general the techniques mentioned herein are well known in the art, reference may be made in particular to Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (1989) and Ausubel *et al.*, Short Protocols in Molecular Biology (1999) 4th Ed. John Wiley & Sons, Inc (as well as the complete version Current Protocols in Molecular Biology).

A. POLYPEPTIDES

It will be understood that polypeptide sequences of the invention are not limited to the particular sequences set forth in SEQ. ID. No. 2 and SEQ. ID. No. 4, or fragments thereof, or sequences obtained from GCR1 or GCR2 protein, but also include homologous sequences obtained from any source, for example related cellular homologues, homologues from other species and variants or derivatives thereof.

Thus, the present invention encompasses variants, homologues or derivatives of the amino acid sequences set forth in SEQ. ID. No. 2 and SEQ. ID. No. 4, as well as variants, homologues or derivatives of the amino acid sequences encoded by the nucleotide sequences of the present invention.

In the context of the present invention, a homologous sequence is taken to include an amino acid sequence which is at least 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical at the amino acid level over at least 30, preferably 50, 70, 90 or 100 amino acids with GCR1 or GCR2, for example as shown in the sequence listing herein.

Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

% homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues (for example less than 50 contiguous amino acids).

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the

GCG Wisconsin Bestfit package (see below) the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux *et al.*, 1984, Nucleic Acids Research 12:387). Examples of other software than can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel *et al.*, 1999 *ibid* – Chapter 18). FASTA (Atschul *et al.*, 1990, J. Mol. Biol., 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel *et al.*, 1999 *ibid*, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program.

Although the final % homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). It is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

B. POLYPEPTIDE VARIANTS, DERIVATIVES AND FRAGMENTS

The terms "variant" or "derivative" in relation to the amino acid sequences of the present invention includes any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) amino acids from or to the sequence. Preferably, nucleic acids according to the invention are understood to comprise variants or derivatives thereof.

In any case, however, the key feature of the sequences of the invention – namely that they are specific for PGCs and can serve as a marker for PGCs in a cell population – is retained.

Natural variants of GCR1 and GCR2 are likely to comprise conservative amino acid substitutions. Conservative substitutions may be defined, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

Polypeptides of the invention useful as markers also include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ. ID. No. 2 and SEQ. ID. No. 4. Preferred fragments include those which include an epitope. Suitable fragments will be at least about 5, e.g. 10, 12, 15 or 20 amino acids in length. They may also be less than 100, 75 or 50 amino acids in length. Polypeptide fragments of the GCR proteins and allelic and species variants thereof may contain one or more (e.g. 5, 10, 15, or 20) substitutions, deletions or insertions, including conserved substitutions. Where substitutions, deletion and/or insertions occur, for example in different species, preferably less than 50%, 40% or 20% of the amino acid residues depicted in the sequence listings are altered.

C. NUCLEIC ACIDS

Polynucleotides according to the invention comprise nucleic acid sequences encoding the polypeptide sequences of the invention. It will be understood by a skilled person that numerous different polynucleotides can encode the same polypeptide as a result of the degeneracy of the genetic code. In addition, it is to be understood that skilled persons may, using routine techniques, make nucleotide substitutions that do not affect the polypeptide sequence encoded by the polynucleotides of the invention to reflect the codon usage of any particular host organism in which the polypeptides of the invention are to be expressed.

Polynucleotides of the invention may comprise DNA or RNA. They may be single-stranded or double-stranded. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or life span of polynucleotides of the invention.

The terms "variant", "homologue" or "derivative" in relation to the nucleotide sequence of the present invention include any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) nucleotides from or to the sequence providing the resultant nucleotide sequence is specific for PGCs.

As indicated above, with respect to sequence homology, preferably there is at least 75%, more preferably at least 85%, more preferably at least 90% homology to the sequences shown in the sequence listing herein. More preferably there is at least 95%, more preferably at least 98%, homology. Nucleotide homology comparisons may be conducted as described above. A preferred sequence comparison program is the GCG Wisconsin Bestfit program described above. The default scoring matrix has a match value of 10 for each identical

nucleotide and -9 for each mismatch. The default gap creation penalty is -50 and the default gap extension penalty is -3 for each nucleotide.

The present invention also encompasses nucleotide sequences that are capable of hybridising selectively to the sequences presented herein, or any variant, fragment or derivative thereof, or to the complement of any of the above. Nucleotide sequences are preferably at least 15 nucleotides in length, more preferably at least 20, 30, 40 or 50 nucleotides in length.

The term "hybridisation" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" as well as the process of amplification as carried out in polymerase chain reaction technologies.

Polynucleotides of the invention capable of selectively hybridising to the nucleotide sequences presented herein, or to their complement, will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% or 98% homologous to the corresponding nucleotide sequences presented herein over a region of at least 20, preferably at least 25 or 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

The term "selectively hybridisable" means that the polynucleotide used as a probe is used under conditions where a target polynucleotide of the invention is found to hybridize to the probe at a level significantly above background. The background hybridization may occur because of other polynucleotides present, for example, in the cDNA or genomic DNA library being screening. In this event, background implies a level of signal generated by interaction between the probe and a non-specific DNA member of the library which is less than 10 fold, preferably less than 100 fold as intense as the specific interaction observed with the target DNA. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P.

Hybridisation conditions are based on the melting temperature (Tm) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Maximum stringency typically occurs at about Tm-5°C (5°C below the Tm of the probe); high stringency at about 5°C to 10°C below Tm; intermediate stringency at about 10°C to 20°C below Tm; and low stringency at about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a maximum stringency hybridisation can be used to identify or detect identical polynucleotide sequences while an intermediate (or low) stringency hybridisation can be used to identify or detect similar or related polynucleotide sequences.

In a preferred aspect, the present invention covers nucleotide sequences that can hybridise to the nucleotide sequence of the present invention under stringent conditions (e.g. 65°C and 0.1xSSC {1xSSC = 0.15 M NaCl, 0.015 M Na_3 Citrate pH 7.0}).

Where the polynucleotide of the invention is double-stranded, both strands of the duplex, either individually or in combination, are encompassed by the present invention. Where the polynucleotide is single-stranded, it is to be understood that the complementary sequence of that polynucleotide is also included within the scope of the present invention.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the scope of the invention can be obtained in a number of ways. Other variants of the sequences described herein may be obtained for example by probing DNA libraries made from a range of individuals, for example individuals from different populations. In addition, other viral/bacterial, or cellular homologues particularly cellular homologues found in mammalian cells (e.g. rat, mouse, bovine and primate cells, including human cells), may be obtained and such homologues and fragments thereof in general will be capable of selectively hybridising to the sequences shown in the sequence listing herein. Such sequences may be obtained by probing cDNA libraries made from or genomic DNA libraries from other animal species, and probing such libraries with probes comprising all or part of SEQ I.D. Nos 1 or 3 under conditions of medium to high stringency. Similar considerations apply to obtaining species homologues and allelic variants of GCR1 and GCR2.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label

by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein. Preferred fragments are less than 500, 200, 100, 50 or 20 nucleotides in length.

Polynucleotides such as a DNA polynucleotides and probes according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15 to 30 nucleotides) flanking a region of the sequence which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from an animal or human cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector

D. NUCLEOTIDE VECTORS

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells include bacteria such as *E. coli*, yeast, mammalian cell

lines and other eukaryotic cell lines, for example insect Sf9 cells.

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence that is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably linked" means that the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

The control sequences may be modified, for example by the addition of further transcriptional regulatory elements to make the level of transcription directed by the control sequences more responsive to transcriptional modulators.

Vectors of the invention may be transformed or transfected into a suitable host cell as described below to provide for expression of a protein of the invention. This process may comprise culturing a host cell transformed with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the protein, and optionally recovering the expressed protein.

The vectors may be for example, plasmid or virus vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used, for example, to transfect or transform a host cell.

Control sequences operably linked to sequences encoding the protein of the invention include promoters/enhancers and other expression regulation signals. These control sequences may be selected to be compatible with the host cell for which the expression vector is designed to be used in. The term "promoter" is well-known in the art and encompasses nucleic acid regions ranging in size and complexity from minimal promoters to promoters including upstream elements and enhancers.

The promoter is typically selected from promoters which are functional in mammalian cells, although prokaryotic promoters and promoters functional in other eukaryotic cells may be used. The promoter is typically derived from promoter sequences of viral or eukaryotic genes. For example, it may be a promoter derived from the genome of a cell in which expression is to occur. With respect to eukaryotic promoters, they may be promoters that function in a ubiquitous manner (such as promoters of α-actin, β-actin, tubulin) or, alternatively, a tissue-specific manner (such as promoters of the genes for pyruvate kinase). They may also be promoters that respond to specific stimuli, for example promoters that bind steroid hormone receptors. Viral promoters may also be used, for example the Moloney murine leukaemia virus long terminal repeat (MMLV LTR) promoter, the Rous sarcoma virus (RSV) LTR promoter or the human cytomegalovirus (CMV) IE promoter.

It may also be advantageous for the promoters to be inducible so that the levels of expression of the heterologous gene can be regulated during the life-time of the cell. Inducible means that the levels of expression obtained using the promoter can be regulated.

In addition, any of these promoters may be modified by the addition of further regulatory sequences, for example enhancer sequences. Chimeric promoters may also be used comprising sequence elements from two or more different promoters described above.

E. HOST CELLS

Vectors and polynucleotides of the invention may be introduced into host cells for the purpose of replicating the vectors/polynucleotides and/or expressing the proteins of the invention encoded by the polynucleotides of the invention. Although the proteins of the invention may be produced using prokaryotic cells as host cells, it is preferred to use eukaryotic cells, for example yeast, insect or mammalian cells, in particular mammalian cells.

Vectors/polynucleotides of the invention may introduced into suitable host cells using a

variety of techniques known in the art, such as transfection, transformation and electroporation. Where vectors/polynucleotides of the invention are to be administered to animals, several techniques are known in the art, for example infection with recombinant viral vectors such as retroviruses, herpes simplex viruses and adenoviruses, direct injection of nucleic acids and biolistic transformation.

F. PROTEIN EXPRESSION AND PURIFICATION

Host cells comprising polynucleotides of the invention may be used to express proteins of the invention. Host cells may be cultured under suitable conditions which allow expression of the proteins of the invention. Expression of the proteins of the invention may be constitutive such that they are continually produced, or inducible, requiring a stimulus to initiate expression. In the case of inducible expression, protein production can be initiated when required by, for example, addition of an inducer substance to the culture medium, for example dexamethasone or IPTG.

Proteins of the invention can be extracted from host cells by a variety of techniques known in the art, including enzymatic, chemical and/or osmotic lysis and physical disruption.

G. ANTIBODIES

Antibodies, as used herein, refers to complete antibodies or antibody fragments capable of binding to a selected target, and including Fv, ScFv, Fab' and $F(ab')_2$, monoclonal and polyclonal antibodies, engineered antibodies including chimeric, CDR-grafted and humanised antibodies, and artificially selected antibodies produced using phage display or alternative techniques. Small fragments, such as Fv and ScFv, possess advantageous properties for diagnostic and therapeutic applications on account of their small size and consequent superior tissue distribution.

The antibodies according to the invention are especially indicated for the detection of PGCs. Accordingly, they may be altered antibodies comprising an effector protein such as a label. Especially preferred are labels which allow the imaging of the distribution of

the antibody *in vivo* or *in vitro*. Such labels may be radioactive labels or radioopaque labels, such as metal particles, which are readily visualisable within an embryo or a cell mass. Moreover, they may be fluorescent labels or other labels which are visualisable on tissue samples.

Recombinant DNA technology may be used to improve the antibodies of the invention. Thus, chimeric antibodies may be constructed in order to decrease the immunogenicity thereof in diagnostic or therapeutic applications. Moreover, immunogenicity may be minimised by humanising the antibodies by CDR grafting [see European Patent Application 0 239 400 (Winter)] and, optionally, framework modification [EP 0 239 400].

Antibodies according to the invention may be obtained from animal serum, or, in the case of monoclonal antibodies or fragments thereof, produced in cell culture. Recombinant DNA technology may be used to produce the antibodies according to established procedure, in bacterial or preferably mammalian cell culture. The selected cell culture system preferably secretes the antibody product.

Therefore, the present invention includes a process for the production of an antibody according to the invention comprising culturing a host, e.g. E. coli or a mammalian cell, which has been transformed with a hybrid vector comprising an expression cassette comprising a promoter operably linked to a first DNA sequence encoding a signal peptide linked in the proper reading frame to a second DNA sequence encoding said antibody protein, and isolating said protein.

Multiplication of hybridoma cells or mammalian host cells in vitro is carried out in suitable culture media, which are the customary standard culture media, for example Dulbecco's Modified Eagle Medium (DMEM) or RPMI 1640 medium, optionally replenished by a mammalian serum, e.g. foetal calf serum, or trace elements and growth sustaining supplements, e.g. feeder cells such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages, 2-aminoethanol, insulin, transferrin, low density lipoprotein, oleic acid, or the like. Multiplication of host cells which are bacterial cells or yeast cells is likewise carried out in suitable culture media known in the art, for example for bacteria in medium LB, NZCYM, NZYM, NZM, Terrific Broth, SOB, SOC, 2 x YT,

or M9 Minimal Medium, and for yeast in medium YPD, YEPD, Minimal Medium, or Complete Minimal Dropout Medium.

In vitro production provides relatively pure antibody preparations and allows scale-up to give large amounts of the desired antibodies. Techniques for bacterial cell, yeast or mammalian cell cultivation are known in the art and include homogeneous suspension culture, e.g. in an airlift reactor or in a continuous stirrer reactor, or immobilised or entrapped cell culture, e.g. in hollow fibres, microcapsules, on agarose microbeads or ceramic cartridges.

Large quantities of the desired antibodies can also be obtained by multiplying mammalian cells in vivo. For this purpose, hybridoma cells producing the desired antibodies are injected into histocompatible mammals to cause growth of antibody-producing tumours. Optionally, the animals are primed with a hydrocarbon, especially mineral oils such as pristane (tetramethyl-pentadecane), prior to the injection. After one to three weeks, the antibodies are isolated from the body fluids of those mammals. For example, hybridoma cells obtained by fusion of suitable myeloma cells with antibody-producing spleen cells from Balb/c mice, or transfected cells derived from hybridoma cell line Sp2/0 that produce the desired antibodies are injected intraperitoneally into Balb/c mice optionally pre-treated with pristane, and, after one to two weeks, ascitic fluid is taken from the animals.

The foregoing, and other, techniques are discussed in, for example, Kohler and Milstein, (1975) Nature 256:495-497; US 4,376,110; Harlow and Lane, Antibodies: a Laboratory Manual, (1988) Cold Spring Harbor, incorporated herein by reference. Techniques for the preparation of recombinant antibody molecules is described in the above references and also in, for example, EP 0623679; EP 0368684 and EP 0436597, which are incorporated herein by reference.

The cell culture supernatants are screened for the desired antibodies, preferentially by immunofluorescent staining of PGCs by immunoblotting, by an enzyme immunoassay, e.g. a sandwich assay or a dot-assay, or a radioimmunoassay.

For isolation of the antibodies, the immunoglobulins in the culture supernatants or in the ascitic fluid may be concentrated, e.g. by precipitation with ammonium sulphate, dialysis against hygroscopic material such as polyethylene glycol, filtration through selective membranes, or the like. If necessary and/or desired, the antibodies are purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose and/or (immuno-) affinity chromatography, e.g. affinity chromatography with GCR1 or GCR2, or fragments thereof, or with Protein-A.

The invention further concerns hybridoma cells secreting the monoclonal antibodies of the invention. The preferred hybridoma cells of the invention are genetically stable, secrete monoclonal antibodies of the invention of the desired specificity and can be activated from deep-frozen cultures by thawing and recloning.

The invention also concerns a process for the preparation of a hybridoma cell line secreting monoclonal antibodies directed to GCR1 and/or GCR2, characterised in that a suitable mammal, for example a Balb/c mouse, is immunised with a one or more GCR1 or GCR2 polypeptides, or antigenic fragments thereof; antibody-producing cells of the immunised mammal are fused with cells of a suitable myeloma cell line, the hybrid cells obtained in the fusion are cloned, and cell clones secreting the desired antibodies are selected. For example spleen cells of Balb/c mice immunised with GCR1 and/or GCR2 are fused with cells of the mycloma cell line PAI or the mycloma cell line Sp2/0-Ag14, the obtained hybrid cells are screened for secretion of the desired antibodies, and positive hybridoma cells are cloned.

Preferred is a process for the preparation of a hybridoma cell line, characterised in that Balb/c mice are immunised by injecting subcutaneously and/or intraperitoneally between 10 and 10^7 and 10^8 cells expressing GCR1 and/or GCR2 and a suitable adjuvant several times, e.g. four to six times, over several months, e.g. between two and four months, and spleen cells from the immunised mice are taken two to four days after the last injection and fused with cells of the myeloma cell line PAI in the presence of a fusion promoter, preferably polyethylene glycol. Preferably the myeloma cells are fused with a three- to twentyfold excess of spleen cells from the immunised mice in a solution containing about

30 % to about 50 % polyethylene glycol of a molecular weight around 4000. After the fusion the cells are expanded in suitable culture media as described hereinbefore, supplemented with a selection medium, for example HAT medium, at regular intervals in order to prevent normal myeloma cells from overgrowing the desired hybridoma cells.

The invention also concerns recombinant DNAs comprising an insert coding for a heavy chain variable domain and/or for a light chain variable domain of antibodies directed to GCR1 and/or GCR2 as described hereinbefore. By definition such DNAs comprise coding single stranded DNAs, double stranded DNAs consisting of said coding DNAs and of complementary DNAs thereto, or these complementary (single stranded) DNAs themselves.

Furthermore, DNA encoding a heavy chain variable domain and/or for a light chain variable domain of antibodies directed to GCR1 and/or GCR2 can be enzymatically or chemically synthesised DNA having the authentic DNA sequence coding for a heavy chain variable domain and/or for the light chain variable domain, or a mutant thereof. A mutant of the authentic DNA is a DNA encoding a heavy chain variable domain and/or a light chain variable domain of the above-mentioned antibodies in which one or more amino acids are deleted or exchanged with one or more other amino acids. Preferably said modification(s) are outside the CDRs of the heavy chain variable domain and/or of the light chain variable domain of the antibody. Such a mutant DNA is also intended to be a silent mutant wherein one or more nucleotides are replaced by other nucleotides with the new codons coding for the same amino acid(s). Such a mutant sequence is also a degenerated sequence. Degenerated sequences are degenerated within the meaning of the genetic code in that an unlimited number of nucleotides are replaced by other nucleotides without resulting in a change of the amino acid sequence originally encoded. Such degenerated sequences may be useful due to their different restriction sites and/or frequency of particular codons which are preferred by the specific host, particularly *E. coli*, to obtain an optimal expression of the heavy chain murine variable domain and/or a light chain murine variable domain.

The term mutant is intended to include a DNA mutant obtained by in vitro mutagenesis of the authentic DNA according to methods known in the art.

For the assembly of complete tetrameric immunoglobulin molecules and the expression of chimeric antibodies, the recombinant DNA inserts coding for heavy and light chain variable domains are fused with the corresponding DNAs coding for heavy and light chain constant domains, then transferred into appropriate host cells, for example after incorporation into hybrid vectors.

The invention therefore also concerns recombinant DNAs comprising an insert coding for a heavy chain murine variable domain of an antibody directed to GCR1 and/or GCR2 fused to a human constant domain g, for example γ1, γ2, γ3 or γ4, preferably γ1 or γ4. Likewise the invention concerns recombinant DNAs comprising an insert coding for a light chain murine variable domain of an antibody directed to GCR1 and/or GCR2 fused to a human constant domain κ or λ, preferably κ.

In another embodiment the invention pertains to recombinant DNAs coding for a recombinant polypeptide wherein the heavy chain variable domain and the light chain variable domain are linked by way of a spacer group, optionally comprising a signal sequence facilitating the processing of the antibody in the host cell and/or a DNA coding for a peptide facilitating the purification of the antibody and/or a cleavage site and/or a peptide spacer and/or an effector molecule.

The DNA coding for an effector molecule is intended to be a DNA coding for the effector molecules useful in diagnostic or therapeutic applications. Thus, effector molecules which are toxins or enzymes, especially enzymes capable of catalysing the activation of prodrugs, are particularly indicated. The DNA encoding such an effector molecule has the sequence of a naturally occurring enzyme or toxin encoding DNA, or a mutant thereof, and can be prepared by methods well known in the art.

H. DETECTION OF PGCs IN CELL POPULATIONS

Polynucleotide probes or antibodies according to the invention may be used for the detection of PGCs in cell populations. As used herein, a "cell population" is any

collection of cells which may contain one or more PGCs. Preferably, the collection of cells does not consist solely of PGCs, but comprises at least one other cell type.

Cell populations according to the invention therefore comprise embryos and embryo tissue, but also adult tissues and tissues grown in culture and cell preparations derived from any of the foregoing.

Polynucleotides according to the invention may be used for detection of GCR1 and GCR2 transcripts in PGCs by nucleic acid hybridisation techniques. Such techniques include PCR, in which primers are hybridised to GCR1 and/or GCR2 transcripts and used to amplify the transcripts, to provide a detectable signal; and hybridisation of labelled probes, in which probes specific for an unique sequence in the GCR1 and/or GCR2 transcript are used to detect the transcript in the target cells.

As noted hereinbefore, probes may be labelled with radioactive, radioopaque, fluorescent or other labels, as is known in the art.

Antibodies according to the invention may also be used to detect GCR1 and/or GCR2. GRC1, in particular, possesses an extracellular domain which may be targeted by an anti-GCR1 antibody and detected at the cell surface. Alternatively, intracellular scFv may be used to detect GCR1 and/or GCR2 within the cell.

Particularly indicated are immunostaining and FACS techniques. Suitable fluorophores are known in the art, and include chemical fluorophores and fluorescent polypeptides, such as GFP and mutants thereof (see WO 97/28261). Chemical fluorophores may be attached to immunoglobulin molecules by incorporating binding sites therefor into the immunoglobulin molecule during the synthesis thereof.

Preferably, the fluorophore is a fluorescent protein, which is advantageously GFP or a mutant thereof. GFP and its mutants may be synthesised together with the immunoglobulin or target molecule by expression therewith as a fusion polypeptide, according to methods well known in the art. For example, a transcription unit may be constructed as an in-frame fusion of the desired GFP and the immunoglobulin or target,

and inserted into a vector as described above, using conventional PCR cloning and ligation techniques.

Antibodies may be labelled with any label capable of generating a signal. The signal may be any detectable signal, such as the induction of the expression of a detectable gene product. Examples of detectable gene products include bioluminescent polypeptides, such as luciferase and GFP, polypeptides detectable by specific assays, such as β-galactosidase and CAT, and polypeptides which modulate the growth characteristics of the host cell, such as enzymes required for metabolism such as HIS3, or antibiotic resistance genes such as G418. In a preferred aspect of the invention, the signal is detectable at the cell surface. For example, the signal may be a luminescent or fluorescent signal, which is detectable from outside the cell and allows cell sorting by FACS or other optical sorting techniques.

Preferred is the use of optical immunosensor technology, based on optical detection of fluorescently-labelled antibodies. Immunosensors are biochemical detectors comprising an antigen or antibody species coupled to a signal transducer which detects the binding of the complementary species (Rabbany *et al.*, 1994 *Crit Rev Biomed Eng* **22**:307-346; Morgan *et al.*, 1996 *Clin Chem* **42**:193-209). Examples of such complementary species include the antigen Zif 268 and the anti-Zif 268 antibody. Immunosensors produce a quantitative measure of the amount of antibody, antigen or hapten present in a complex sample such as serum or whole blood (Robinson 1991 *Biosens Bioelectron* **6**:183-191). The sensitivity of immunosensors makes them ideal for situations requiring speed and accuracy (Rabbany *et al.*, 1994 *Crit Rev Biomed Eng* **22**:307-346).

Detection techniques employed by immunosensors include electrochemical, piezoelectric or optical detection of the immunointeraction (Ghindilis *et al.*, 1998 *Biosens Bioelectron* **1**:113-131). An indirect immunosensor uses a separate labelled species that is detected after binding by, for example, fluorescence or luminescence (Morgan *et al.*, 1996 *Clin Chem* **42**:193-209). Direct immunosensors detect the binding by a change in potential difference, current, resistance, mass, heat or optical properties (Morgan *et al.*, 1996 *Clin Chem* **42**:193-209). Indirect immunosensors may encounter fewer problems due to non-specific binding (Attridge *et al.*, 1991 *Biosens Bioelecton* **6**:201-214; Morgan *et al.*, 1996 *Clin Chem* **42**:193-209).

Examples

Example 1

Identification of genes specific to the earliest population of primordial germ cells (PGCs) by single cell cDNA differential screening

A method for single cell analysis is developed to identify genes that are involved in the specification of the germ cell lineage, which results in the establishment of a founder population of Primordial Germ Cells (PGCs). It is determined that the lineage specification of PGCs accompanies the expression of a unique set of genes, which are not expressed in somatic cells.

The method for the identification of the genes is mainly based on the differential screening of the libraries made from single cells from day 7.25 mouse embryonic fragments that contain PGCs. The single cell cDNA differential screen was originally described by Brady and Iscove (1), and subsequently modified by Cathaline Dulac and Richard Axel which resulted in the successful identification of the pheromone receptor genes from rat (2). The method of Axel's group is employed, with slight modifications as described.

Construction of single cell cDNAs from embryonic fragment bearing the earliest population of PGCs

In the mouse, the earliest population of the PGCs is reported to consist of alkaline phosphatase positive cluster of some 40 cells, at the base of the emerging allantois at day 7.25 of gestation (3). The precise location of the PGC cluster in the inbred 129Sv and C57BL/6 strain is determined by microscopy using both whole-mount alkaline phosphatase staining and semi-thin sections stained by methylene blue. The earliest stage at which a cluster of PGCs can be detected is at the Late Streak stage (4), when a distinctively stained population of cells is found just beneath an epithelial lining from which the allantoic bud appears. This region is at the border between the extraembryonic

and embryonic tissues just posterior to and above the most proximal part of the primitive streak. The cluster persists at this position at least until Early/Mid Bud stage. In the inbred 129Sv strain, the PGC cluster is found to contain a slightly larger number of the cells, which are more tightly packaged than in the C57BL/6 strain. The 129Sv strain is used for subsequent experiments, as a better recovery of the earliest PGCs is obtained.

129Sv embryos are isolated at E7.5 in DMEM plus 10% FCS buffered with 25mM HEPES at room temperature and the developmental stage of each embryo is determined under a dissection microscope. The precise developmental stage can differ substantially even amongst embryos within the same litter. Embryos that are at the no bud or early bud (allantoic) stage are chosen for further dissection, which in part is dictated by the ease of identification of the region containing PGCs as seen under the dissection microscope. The fragment that is expected to contain the PGC cluster is cut out very precisely by means of solid glass needles. This region is dissociated it into single cells using 0.25% trypsin-1mM EGTA/PBS treatment at 37°C for 10 min, followed by gentle pipetting with a mouth pipette. The dissected fragment usually contained between 250-300 cells. The procedure for cell dispersal with this gentle procedure left the visceral endoderm layer remained as an intact cellular sheet.

We picked single cells randomly from the cell suspension by a mouth pipette and put individual single cells (but avoiding generating air bubbles), into a thin-walled PCR tube containing 4µl of ice-cold cell lysis buffer (50mM Tris-HCl pH8.3, 75mM KCl, 3mM $MgCl_2$, 0.5% NP-40, containing 80ng/ml pd(T)24, 5µg/ml prime RNase inhibitor, 324U/ml RNA guard, and 10mM each of dATP, dCTP, dGTP, and dTTP). The volume of medium carried with the single cell is less than 0.5µl. The tube is briefly centrifuged to ensure that the cell is indeed in the lysis buffer. During each separate experiment, we picked a total of 19 single cells, and left one tube without a cell, to serve as a negative control for the PCR amplification procedure. All the cells that are collected in tubes are kept on ice before starting the subsequent procedure.

The cells are lysed by incubating the tubes at 65°C for 1min, and then kept at room temperature for 1-2 min to allow the oligo dT to anneal the to RNA. First-strand cDNA synthesis is initiated by adding 50U of Moloney murine leukaemia virus (MMLV) and

0.5U of avian mycloblastosis virus (AMV) reverse transcriptase followed by incubation for 15min at 37°C. The reverse transcriptases are inactivated for 10min at 65°C. This reverse transcription reaction is restricted to 15 min. which allows the synthesis of relatively uniform size cDNAs of between 500 base -1000 bases in length from the C termini. This enables the subsequent PCR amplification to be fairly representative.

Next, in order to add the poly A tail to the 5 prime end of the synthesised first-strand cDNA, 4.5µl of 2X tailing buffer (200mM potassium cacodylate pH7.2, 4mM $CoCl_2$, 0.4mM DTT, 200mM dATP containing 10U of terminal transferase) is added to the reaction followed by incubation for 15min at 37 °C. The samples are heat inactivated for 10 min at 65°C. The reaction now contained synthesised cDNAs bearing poly T tail at their C termini and poly A stretch at their N termini, ready for the amplification by the PCR using the specific primer.

The contents of each tube is brought to 100µl with a solution made of 10mM Tris-HCl pH8.3, 50mM KCl, 2.5mM $MgCl_2$, 100µg/ml bovine serum albumin, 0.05% Triton-X 100, 1mM of dATP, dCTP, dGTP, dTTP, 10U of Taq polymerase, and 5µg of the AL1 primer. The AL1 sequence is ATT GGA TCC AGG CCG CTC TGG ACA AAA TAT GAA TCC $(T)_{24}$. The PCR amplification is performed according to the following schedule: 94°C for 1 min, 42°C for 2 min, and 72°C for 6 min with 10 s extension per cycle for 25 cycles. Five additional units of Taq polymerase are added before performing 25 more cycles with the same programme but without the extension time. Each tube at this point contains amplified cDNA products derived from a single cell. The protein contents of the solution are extracted by phenol/chloroform treatment, and the amplified cDNAs are precipitated by ethanol and eventually suspended in 100µl of TE pH8.0. 5µl of the cDNA solution is run on a 1.5% agarose gel to check the success of the amplification. Most of the samples show a very intense 'smeared' band ranging mainly between 500bp to 1200bp, indicating the efficient amplification of the single cell cDNA. Only the successfully amplified samples are used for the subsequent 'cell typing' analysis.

Example 2

Identification of PGCs by examination of the expression of marker genes

The embryonic fragment which is excised theoretically contains three major components: the allantoic mesoderm, PGCs, and extraembryonic mesoderm surrounding PGCs. In order to identify the single cell cDNA of PGC origin amongst these samples, positive and negative selection of the constructed cDNAs is performed, by examining the expression of four marker genes (BMP4, TNAP, Hoxb1, and Oct4), which are known to be either expressed or repressed in various cell types in this region.

At the No/Early Bud stage, BMP4 is reported to be expressed in the emerging allantois and mesodermal components of the developing amnion, chorion, and visceral yolk sac (5). The boundary of BMP4 expression is very sharp, and the expression is completely excluded in the mesodermal region beneath the epithelial lining continuous from the amnionic mesoderm where the putative PGCs are determined. Therefore, BMP4 is used as a negative marker for the selection. Primer pairs are designed for amplifying the C terminal portion of BMP4 (5': GCC ATA CCT TGA CCC GCA GAA G, 3': AAA TGG CAC TCA GTT CAG TGG G). The PCR amplification is performed using 0.5μl of the cDNA solution as a template according to the following schedule: 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min for 20 cycles. Among 83 samples tested, 57 samples show the expected size of bands, indicating expression of BMP4 these single cells. These samples are considered to be of allantoic mesodermal origin, and therefore excluded from amongst the candidates representing cells of PGC origin.

The expression of tissue non-specific alkaline phosphatase (TNAP), which has long been used as an early marker for PGCs (3), is then examined. Primer pairs are designed (5': CCC AAA GCA CCT TAT TTT TCT ACC, 3': TTG GCG AGT CTC TGC AAT TGG) and the same PCR reaction as above is performed. Amongst the 26 samples, 22 samples are judged to be positive for TNAP. From the alkaline phosphatase staining of the sectioned embryos, it is known that the somatic cells surrounding PGCs also express some amount of TNAP, although the level of expression is slightly lower than that in

PGCs. Therefore, amongst these 22 positive samples there should be still be cells destined to become somatic cells as well as PGCs.

One of the genes known to be expressed in the totipotent PGCs but not in somatic cells is Oct4 (6). To examine the possibility that Oct4 can be used as a marker to distinguish PGCs from somatic cells at this stage, Oct4 expression is checked in the 22 samples by PCR (5': CAC TCT ACT CAG TCC CTT TTC, 3': TGT GTC CCA GTC TTT ATT TAA G). All the 22 samples express Oct4 at comparable levels, suggesting that the somatic cells at this stage are still actively transcribing Oct4 RNA.

The amount of expression of TNAP is quantitated in 22 samples by Southern blot analysis (reverse northern blot analysis). Given the fairly representative amplification of the single cell method, confirmed by amplifying single ES cell cDNA, Southern blot analysis allows semi-quantitative measurement of the amount of the genes expressed in the original single cells, although it does not serve as a perfect indicator of cell identity. However, as a result of this TNAP analysis, 10 samples out of 22 show relatively stronger bands at an equivalent level, while the remaining 12 samples exhibit weaker signals. These results suggest that these 22 samples can be divided at least into two groups, one with stronger TNAP expression (therefore from putative PGCs) and the other with weaker TNAP.

The possibility that somatic cells surrounding PGCs start to express Hoxb1, while PGCs do not (personal communication from Dr. Kirstie Lawson) is also examined. Primer pairs are designed (5': AAC TCA TCA GAG GTC GAA GGA, 3': CGG TGC TAT TGT AAG GTC TGC) and the same PCR reaction as above is performed. Among the 22 samples tested, 12 are positive, and more importantly, these 12 samples perfectly match the ones which show weaker TNAP signals, by Southern blot analysis.

Taking all these results into consideration, it is concluded that 10 samples out of 83, which are Oct4 (+), TNAP (++), BMP4 (-), and Hoxb1(-),are of PGC origin. This ratio (10/83) is reasonable, considering the number of the founding population of PGCs as 40 and the number of cells in the fragment as 250-300.

Example 3

Differential screening of single cell cDNA libraries

As the efficiency of the amplification of cDNA differs in each tube, it is very important to select the samples with the most efficiently amplified cDNA for the construction of libraries. The amplification of six different genes (ribosomal protein S12, intermediate filament protein vimentin, β tubulin-5, α actin, Oct4, E-cadherin) is examined in the 10 PGC candidate samples, by Southern blot analysis. Judging from the overall profile of the amplification of all these six genes, three cDNA preparations are selected for the construction of libraries.

To obtain the maximum amount of double strand cDNA, an extension step is performed with 5μl of cell cDNA in 100μl of the PCR buffer described as above (including 1μl of Amplitaq) according to the following schedule: 94°C for 5min, 42°C for 5min, 72°C for 30min. The solution is extracted by phenol/chloroform treatment, and the amplified cDNAs are precipitated by ethanol, suspended in TE, and completely digested with EcoRI. The PCR primer and excess amount of dNTPs are removed by QIAGEN PCR Purification Kit, and all the purified cDNAs are run on a 2% low melting agarose gel. cDNAs above 500bp are cut and purified by QIAGEN Gel Purification Kit. The purified cDNAs are precipitated by ethanol and suspended in TE and ligated into λ ZAP II vector arms. The ligated vector is packaged, titered and the ratio of the successfully ligated clones is monitored by amplifying the inserts with T3 and T7 primers from 20 plaques. More than 95% of the phage are found to contain inserts.

The representation of the three genes, ribosomal protein S12, β tubulin-5, Oct4, is quantitated by screening 5000 plaques, and the library of the best quality among the three (S12 0.62%, β tubulin 0.4%, Oct4 0.5%) is used for the differential screening. As a comparison partner with the PGC probe, one of the most efficiently amplified surrounding somatic cell cDNA (Oct4 (+), TNAP(+/-), BMP(-), and Hoxb1(+)) is selected by the similar Southern blot analysis.

The library is plated at a density of 1000 plaques per 15cm dish to obtain large plaques (2mm diameter) and two duplicate lifts are taken using Hybond N+ filters from Amersham. The filters are prehybridized at 65°C in 0.5M sodium phosphate buffer (pH7.3) containing 1% bovine serum albumin and 4% SDS. We prepared the cell cDNA probes by reamplifying for 10 cycles 1µl of the original cell cDNA into 50µl of total reaction with the A1.1 primer, in the absence of cold dCTP and with 100µCi of newly received ^{32}PdCTP, followed by the purification using Amersham Nick™ Spin Column. The filters are hybridised for at least 16 hrs with $1.0X10^7$cpm/ml (The first filter is hybridised with somatic cell probe and the second filter is hybridised with the PGC probe). After the hybridisation, the filters are washed three times at 65°C in 0.5X SSC, 0.5% SDS and exposed to X ray films until the appropriate signal is obtained (usually one to two days).

The positive plaques in the two duplicate filters are compared very carefully. Among 5000 plaques screened, 280 are picked as candidates representing the differentially expressed genes. The inserts of all the 280 plaques are amplified with T3 and T7 primers, run on 1.5% gels, and double sandwich Southern blotted. Each membrane is hybridised with the PGC and somatic cell probe, respectively, using the same conditions as the screening. 38 clones amongst the 280 are selected as differentially expressed genes. These clones are next hybridised with the second PGC and somatic cell cDNA probes, which resulted in 20 clones out of 38 to be common in both PGC cDNAs but they are either not included or less abundant in both somatic cell cDNAs. The sequences of all the 20 clones are determined.

Genes highly specific to the earliest population of PGCs

The 20 clones represent 11 different genes (two clones appear two times, one clone appears three times, and one clone appears 6 times). To further stringently check the specificity of expression, primer pairs are designed for these 11 clones and their expression checked in 10 different single PGC-candidate cDNAs and 10 different single somatic cell cDNAs by PCR. Two of them show highly specific expression to PGC cDNAs.

The first gene, GCR1 (Germ cell restricted-1), encodes a 137 amino acid protein. The best fit model of the EMBL program PredictProtein predicts two transmembrane domains, both N and C terminus ends being located outside. Database searches reveal a sequence match with the rat interferon-inducible protein (sp:INIB RAT, pir:JC1241) with unknown function. The GCR1 sequence appears six times in our screen, indicating high level expression in PGCs.

The second gene, GCR2, encodes a 150 amino acid protein with very basic pI (pI=9.67) and containing a nuclear localisation signal. This protein has no sequence match with any known proteins. EST database searches reveal that this sequence is only found in the preimplantation embryo and germ line (newborn ovary and female 12.5 mesonephros and gonad etc) ESTs, indicating the specificity of the expression in the totipotent or pluripotent cells.

Example 4

Identification of PGCs by screening for GCR1 and GCR2 expression

Although PGCs are identified in Example 2 by analysis of BMP4, TNAP, Hoxb1, and Oct4, no single one of these genes can be taken as a marker for the PGC state. However, both GCR1 and GCR2 may be used as such in the present invention.

The boundary of GCR2 expression in particular is well-defined, and the expression is substantially limited to PGCs. Therefore, GCR2 is used as a positive marker for the selection of PGCs. Primer pairs are designed for amplifying the C terminal portion of GCR2 (5': GCCATTCAGATGTCTCTGCAC, 3': CTCACAGCTTGAGGCTTCTAA). The PCR amplification is performed using 0.5µl of the cDNA solution obtained from PGCs in Example 1 as a template according to the following schedule: 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min for 20 cycles. Among 83 samples tested, only those taken from PGCs show expression of GCR2. Hence, GCR2 is a positive marker for the PGC fate.

The expression of GCR1 is then examined. Primer pairs are designed (5': CTACTCCGTGAAGTCTAGG, 3': AATGAGTGTTACACCTGCGTG) and the same PCR reaction as above is performed. Again, amongst the 85 samples tested, only those previously determined to be derived from PGCs are identified as expressing GCR1.

Accordingly, both GCR1 and GCR2 are positive markers for the PGC fate which can be used to positively identify PGC.

Identification of PGC by ISH

The *in vivo* expression of the two genes is examined by *in situ* hybridisation. The expression of GCR1 starts very weakly in the entire epiblast at E6.0-E6.5 (PreStreak stage) and becomes strong in the few cell layers of the proximal rim of the epiblast. The expression seems to become more intense at the proximo-posterior end of the developing primitive streak at the Early/Mid Streak stage and becomes very strong at this position from Late Streak stage onward. The expression persists until Early Head Fold stage and eventually disappears gradually. No expression is detected in the migrating PGCs at E8.5.

The expression of GCR2 starts at the proximo-posterior end of the developing primitive streak at Mid/Late Streak stage and becomes gradually strong at the same position from the later stage onward. The expression is specific and individual single cells stained in a dotted manner can be seen in the region where PGCs are considered to start differentiating as a cluster of cells. At Late Bud/Early Head Fold stage, some cells considered to be migrating from the initial cluster are stained as well as cells in the cluster. At E8.5 and E9.5, a group of cells considered to be the migrating PGCs are very specifically stained.

From these results, it is concluded that GCR1 is a gene which is upregulated during the process of lineage specification of PGCs, and GCR2 is a gene which is turned on after GCR1 to fix the PGC fate.

References

1. Brady, G. and Iscove, N.N. (1993). Construction of cDNA libraries from single cells. *Methods Enzymol.* 225, 611-623.

2. Dulac, C. and Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell* 83, 195-206.

3. Ginsburg, M., Snow, M.H.L., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* 110, 521-528.

4. Downs, K.M., and Davies, T. (1993). Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* 118, 1255-1266.

5. Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M., and Hogan, B.L.M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes&Dev.* 13, 424-436.

6. Yoem, Y.H., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 1996, 881-894.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Claims

1. A nucleic acid having at least 90% homology with the sequence set forth in SEQ. ID. No. 1.

2. A nucleic acid having at least 75% homology with the sequence set forth in SEQ. ID. No. 3.

3. A nucleic acid comprising a sequence of 25 contiguous nucleotides of the nucleic acid of claim 1 or claim 2.

4. A nucleic acid comprising a sequence of 15 contiguous nucleotides of the nucleic acid of claim 1 or claim 2.

5. The complement of a nucleic acid sequence according to any preceding claim.

6. A nucleic acid according to any one of claims 1 to 5, comprising one or more nucleotide substitutions, wherein such substitutions do not alter the coding specificity of said nucleic acid as a result of the degeneracy of the genetic code.

7. A polypeptide encoded by a nucleic acid according to any preceding claim.

8. A method for identifying a primordial germ cell in a population of cells, comprising detecting the expression of a nucleic acid sequence according to claim 1 or claim 2, or a homologue thereof.

9. A method according to claim 8, comprising the steps of amplifying nucleic acids from putative PGCs using 5' and 3' primers specific for GCR1 and/or GCR2, and detecting amplified nucleic acid thus produced.

10. A method according to claim 8, wherein the expression of the nucleic acid sequence is detected by *in situ* hybridisation.

11. A method according to claim 8, wherein the expression of the nucleic acid sequence is determined by detecting the protein product encoded thereby.

12. A method according to claim 11, wherein the protein product is detected by immunostaining.

13. An antibody specific for a polypeptide according to claim 7.

14. An antibody according to claim 13, specific for the extracellular domain of GCR1.

15. Use of an antibody according to claim 13 or claim 14 for the identification of a PGC in a population of cells.

16. A PGC when identified by a method according to any one of claims 8 to 12.

17. A method for isolating a gene specifically expressed in PGCs, comprising the steps of:

a) providing a population of cells containing PGCs;

b) isolating one or more PGCs therefrom and providing single-cell PGC isolates;

c) amplifying the transcribed nucleic acid present in a single PGC;

d) conducting a subtractive hybridisation screen to identify transcripts present in PGCs but not in somatic cells; and

e) probing a nucleic acid library with one or more transcripts identified in d) to clone one or more genes which are specifically expressed in PGCs.

Abstract

The invention provides two primordial germ cell-specifically expressed genes, GCR1 and GCR2, which are markers for primordial germ cells and may be used to identify such cells in cell populations.

SEQUENCE LISTING

SEQ. ID. No. 1

GCR1 full length

GCCGCAGAAAGGGCAGACCCGCAGCGCGCTCCATCCTTTGCCCTCCAGTGCT
GCCTTTGCTCCGCACCATGAACCACACTTCTCAAGCCTTCATCACCGCTGCCA
GTGGAGGACAGCCCCCAAACTACGAAAGAATCAAGGAAGAATATGAGGTGG
CTGAGATGGGGGCACCGCACGGATCGGCTTCTGTCAGAACTACTGTGATCAA
CATGCCCAGAGAGGTGTCGGTGCCTGACCATGTGGTCTGGTCCCTGTTCAATA
CACTCTTCATGAACTTCTGCTGCCTGGGCTTCATAGCCTATGCCTACTCCGTGA
AGTCTAGGGATCGGAAGATGGTGGGTGATGTGACTGGAGCCCAGGCCTACGC
CTCCACTGCTAAGTGCCTGAACATCAGCACCTTGGTCCTCAGCATCCTGATGG
TTGTTATCACCATTGTTAGTGTCATCATCATTGTTCTTAACGCTCAAAACCTTC
ACACTTAATAGAGGATTCCGACTTCCGGTCCTGAAGTGCTTCACCCTCCGCAG
CTGCGTCCCTCCTTGCCCCTCCCTACACGCAGGTGTAACACTCATTTATCTATC
CACAGTGGATTCAATAAAGTGCACTTGATAACCACC

SEQ. ID. No. 2

GCR2 full length

GGATCACAGACTGACTGCTAATTGGGTCTTGGTTTTAGGTCTTTTCAAAGACT
AAGCAATCTTGTTCCGAGCTAGCTTTTGAGGCTTCTGCCCATCGCATCGCCAT
GGAGGAACCATCAGAGAAAGTCGACCCAATGAAGGACCCTGAAACTCCTCAG
AAGAAAGATGAAGAGGACGCTTTGGATGATACAGACGTCCTACAACCAGAA
ACACTAGTAAAGGTCATGAAAAAGCTAACCCTAAACCCCGGTGTCAAGCGGT
CCGCACGCCGGCGCAGTCTACGGAACCGCATTGCAGCCGTACCTGTGGAGAA
CAAGAGTGAAAAAATCCGGAGGGAAGTTCAAAGCGCCTTTCCCAAGAGAAG
GGTCCGCACTTTGTTGTCGGTGCTGAAAGACCCTATAGCAAAGATGAGAAGA
CTTGTTCGGATTGAGCAGAGACAAAAAAGGCTCGAAGGAAATGAGTTTGAAC
GGGACAGTGAGCCATTCAGATGTCTCTGCACTTTCTGCCATTATCAAAGATGG

GATCCCTCTGAGAATGCGAAAATCGGGAAGAATTAGGAGCTTACATTGTACG
CTGCCCTGGCTGTCGACGATGCCGCACAGCAGATGTGAAAGCTATTTTTTGTT
TAAGATTAAACTTTTTCTGGTGCTGGGAAATCTTAACTTGTTAACCTTTAAATT
GTAGATAGGATGCACAACGATCCAGATTTATGTGAAGTTTAGAAGCCTCAAG
CTGTGAGGCCCAGGGCTGAGGAATAAAGTAAATAGAATTTGGAGTATGTACG
TTCTAATTTCCAGAAATTTGTAATAAAAGCATTTTTGTT